HOW WELL CAN A MUSIC EMOTION RECOGNITION SYSTEM PREDICT THE EMOTIONAL RESPONSES OF PARTICIPANTS?

Yading Song and Simon Dixon
Centre for Digital Music
Queen Mary University of London
{y.song, s.e.dixon}@qmul.ac.uk

ABSTRACT

Music emotion recognition systems have been shown to perform well for musical genres such as film soundtracks and classical music. It seems difficult, however, to reach a satisfactory level of classification accuracy for popular music. Unlike genre, music emotion involves complex interactions between the listener, the music and the situation. Research on MER systems is handicapped due to the lack of empirical studies on emotional responses. In this paper, we present a study of music and emotion using two models of emotion. Participants' responses on 80 music stimuli for the categorical and dimensional model, are compared. In addition, we collect 207 musical excerpts provided by participants for four basic emotion categories (happy, sad, relaxed, and angry). Given that these examples represent intense emotions, we use them to train musical features using support vector machines with different kernels and with random forests. The most accurate classifier, using random forests, is then applied to the 80 stimuli, and the results are compared with participants' responses. The analysis shows similar emotional responses for both models of emotion. Moreover, if the majority of participants agree on the same emotion category, the emotion of the song is also likely to be recognised by our MER system. This indicates that subjectivity in music experience limits the performance of MER systems, and only strongly consistent emotional responses can be predicted.

1. INTRODUCTION

With technological and social changes in our daily lives, the experience of music has changed at a fundamental level. Music can be heard at far more diverse places, and people report that the primary reason for listening to music lies in its emotional effects, the induction and expression of emotions [1]. Because of the emotional function of music, over the past decade, the study of music and emotion has become increasingly important, and has attracted research from different fields, for instance, computer science [2], musicology, and psychology [3]. Previous studies on emotion provide us with a better understanding of music and emotion, which can also help improve the design of subjective music recommendation systems [4].

For music information retrieval (MIR) researchers, music emotion recognition (MER) systems have been widely discussed [2, 5]. On the one hand, previous studies using musical features have applied various machine learning approaches (e.g., support vector machines [6], k-nearest neighbours [7], random forests [8], regression models [9], and deep belief networks [10, 11]). Although these techniques perform well for genres such as classical music and film soundtracks, the recognition accuracy for popular music fails to reach a satisfactory level [7, 8]. On the other hand, psychological studies in music have focussed on emotional responses to music [12], emotion models [13], emotion experience and recognition [14], and cross-cultural emotion perception in music [15, 16]. The comparison between listeners' responses using the categorical and dimensional model are often neglected [13]. Other research also suggests that differences in individuals may affect how emotional meaning is elicited [17, 18]. In this study, however, we compare participants' responses in general, rather than individual factors such as one's personality, current mood, or culture.

Emerging from research in both computer science and psychology, we study the differences between music emotion recognition systems and participants' responses using two models of emotion, the categorical and dimensional model. Therefore, the goals of this paper are, (1) To compare participants' responses for two models of emotion; (2) To provide a user-suggested dataset of musical excerpts for four basic emotions (happy, sad, relaxed and angry); (3) To study the differences between machine learning approaches (e.g., support vector machines and random forest) and participants' responses.

2. MUSIC AND EMOTION

2.1 Emotional Responses

One important distinction in music is between *perceived emotion* (or expressed emotion) which is an emotion expressed by music, and *induced emotion* (or felt emotion) which is an emotion felt in response to music. In general, music evokes emotions similar to the emotions perceived in music [18,19]. However, some research suggests that responses for induced emotion are generally positive [1], and responses for perceived emotion are more consistent [14].

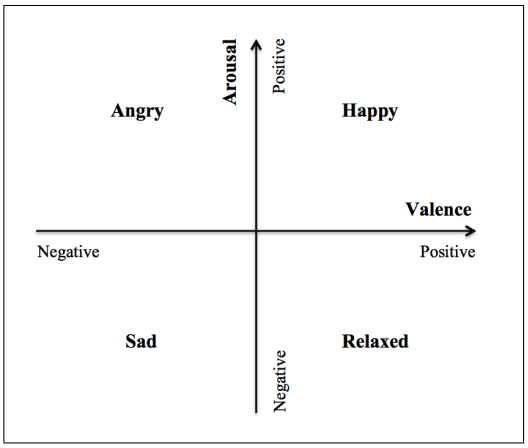

Figure 1. A mapping between a categorical (happy, sad, relaxed and angry) and dimensional model (valence and arousal) of emotion.

2.2 Emotion Models

Although different emotion models such as miscellaneous [20] and domain-specific [21] models have been proposed in the past, the most popular ones are the categorical and dimensional models. The typical dimensional models of emotion represent emotions in an affective space with two dimensions: one related to valence (a pleasure-displeasure continuum), and the other to arousal (activation-deactivation) [22]. Previous studies using the dimensional model have suggested that prediction for arousal is more consistent than for valence [23]. In contrast, the categorical model represents all emotions as being derived from a limited number of universal and innate basic emotions such as happiness, sadness, fear, and anger [24]; and is often used in the study of perceived emotion [3]. In this study, both categorical and dimensional models are used, and to compare the results between these two models of emotion, a mapping is provided in Figure 1. We used four basic emotion classes: happy, angry, sad, and relaxed, considering these four emotions are widely accepted across different cultures and cover the four quadrants of the two-dimensional model of emotion [25].

3. DATA COLLECTION

The majority of studies on music and emotion have used film soundtracks and classical music [17, 26]. Compared with other musical genres, there has been a lack of MER research on popular music [25, 27–29]. Although social tags provide us with highly relevant metadata such as genre, mood, and instrument [30], participants' agreement with emotion tags such as "relaxed" is still very low [31, 32].

3.1 Musical Excerpt Collection

In a previous listening experiment on music emotion using the categorical model [32], forty participants were asked to provide examples of songs (song title and artist's name) that represent each of the four basic emotions (happy, sad, relaxed, and angry) in perceived and induced emotion. Given that music evokes emotions similar to the emotions perceived in music, the examples for perceived and induced emotion were aggregated for this study. If the same excerpt was mentioned in both perceived and induced emotion, the song is only counted once. However, some participants mentioned only the artist name (e.g., Death Cab for Cutie, Mayday Parade, and Bandari), or the album name (e.g., The Dark Side of the Moon), so this information was not considered for further analysis. Musical excerpts were then fetched via the 7Digital API [1] or Amazon mp3 [2]. A total of 207 songs were collected in this way, with the distribution over emotion categories as shown in Table 1.

In contrast to songs retrieved using emotion tags, these examples are considered more likely to represent intense emotions. A music example from each emotion category is shown in Table 2. The dataset (song title, artist's name, 7digital ID, and musical features) is made available to encourage other researchers to reproduce the results for research and evaluation [3].

Emotion category	No. of examples
Happy	59
Sad	58
Relaxed	48
Angry	42
Total	207

Table 1. The distribution of musical examples provided by participants.

Emotion category	Song title	Artist name
Happy	Wannabe	Spice Girls
Sad	Fix You	Coldplay
Relaxed	Eggplant	Michael Franks
Angry	Fighter	Christina Aguilera

Table 2. User-provided examples for each emotion category.

3.2 Emotion Ratings

A separate eighty (n = 20 for each emotion category) popular musical excerpts were randomly selected from a data set of 2904 songs that had been tagged with one of the four words "happy", "sad", "relaxed", and "angry" [6]. These 80 musical excerpts were given in random order to forty participants using the categorical model [31], and fifty-four participants using the dimensional model [32]. Previous research showed a higher consistency in participants' perceived emotional responses. Therefore, only perceived emotional responses are considered in this study. For the categorical model, participants were asked to choose from one of the following options: happy, sad, relaxed, angry, and "cannot tell"/"none of the above". For the dimensional model, participants were asked to rate on an 11-point scale for the two core dimensions: valence (sad-happy) and arousal (calm-excited). Their ratings were aggregated, and a summary of the responses, participants' profiles, and musical excerpts is made publicly available [19] [3].

[1] http://developer.7digital.com/

[2] http://www.amazon.co.uk/Digital-Music/b?ie=UTF8&node=77197031

[3] https://code.soundsoftware.ac.uk/projects/emotion-recognition/repository

Dimension	Description
Dynamics	RMS energy, slope, attack, low energy
Rhythm	tempo, fluctuation peak (pos, mag)
Spectral	spectrum centroid, brightness, spread, skewness, kurtosis, rolloff95, rolloff85, spectral energy, spectral entropy, flatness, roughness, irregularity, zero crossing rate, spectral flux, MFCC, DMFCC, DDMFCC
Harmony	chromagram peak, chromagram centroid, key clarity, key mode, HCDF

Note. The mean and standard deviation values were extracted, except for the feature "low energy", for which only the mean was calculated.

Table 3. Features extracted from the audio data.

3.3 Musical Feature Extraction

Two different emotion datasets, training and testing, are used in our experiment. The training dataset, which is provided by participants, contains 207 songs. The testing dataset contains 80 musical excerpts (*n* = 20 for each emotion category). These musical excerpts for testing range from recent releases back to 1960s, and cover a range of Western popular music styles such as pop, rock, country, metal, and instrumental. Each excerpt was either 30 seconds or 60 seconds long (as provided by 7Digital [1]). Previous studies have suggested that emotion can be recognised within a second [17,33]. To expand both the training and testing datasets, each excerpt was split into 5-second clips with 2.5-second overlap. Musical features were then extracted using MIRtoolbox 1.5 [34] [4] for both the full 30/60-second excerpts and the 5-second clips. The musical features extracted are shown in Table 3.

4. RESULTS

4.1 Participants' Responses for the Two Models of Emotion

To compare participants' responses for the categorical and dimensional models, their ratings were aggregated by label (for the categorical model: happy, sad, relaxed, and angry; for the dimensional model: valence and arousal). The ratings of valence and arousal in the dimensional model were mapped to the four basic emotions in the categorical model (see Figure 1). We calculated the inter-rater reliability (Fleiss's Kappa) for participants' ratings using the categorical ($\kappa = 0.31$) and dimensional model ($\kappa = 0.25$). In addition, for each stimulus, we took the label with the greatest number of votes to be the dominant emotion in each model. If the same dominant emotion was found in both categorical and dimensional models, the song was marked as a "match" (53 cases), otherwise "no match" (27 cases). However, three responses using the categorical model and one response using the dimensional model received equal number of votes (e.g., angry with happy, happy with sad, and sad with relaxed), and they were considered as "no match".

Although over the half of the excerpts received the same emotion in both models of emotion, the participants' consistency (the greatest number of votes on the four emotions) between "match" and "no match" cases is still unclear. Therefore, a Kruskal-Wallis one-way analysis of variance test was conducted on participants' consistency between "match" and "no match" cases for two models of emotion. As expected, a significant higher consistency can be found for "match" cases in both categorical ($Median = 0.70$, $Std = 0.19$, $\chi^2(1, N = 80) = 8.79$, and $p < .05$) and dimensional models ($Median = 0.74$, $Std = 0.09$, $\chi^2(1, N = 80) = 4.91$, and $p < .05$) than "no match" cases (for categorical model: $Median = 0.50$, and $Std = 0.13$; for dimensional model:$Median = 0.68$, and $Std = 0.10$). However, no significant differences were found for the two core dimensions, valence ($\chi^2(1, N = 80) = 3.79$, and $p > .05$) and arousal ($\chi^2(1, N = 80) = 1.05$, and $p > 0.05$).

Among the 27 "no match" cases, 10 were collectively confused between the emotions "sad" and "relaxed". When participants' responses were not consistent, it is important to know how machine learning approaches perform. Therefore, we built emotion classifiers using support vector machines and random forest approaches.

4.2 Emotion Recognition Using Machine Learning Approaches

207 excerpts provided by participants were used for training (see Section 3.3). However, a smaller training size may influence classification performance. To expand the data, each audio file was split into 5-second clips with 2.5-second overlap. Therefore, 207 (30/60 seconds) and 2990 (5 seconds) musical clips were collected, and trained separately.

4.2.1 Training

We adopted a 10 fold cross-validation approach, where for each song, all clips were placed in a single fold to avoid overfitting, and chose support vector machines (SVM) with different kernels (e.g., linear, radial basis function, and polynomial) and random forests (RF) as classifiers for training. We used the implementation of the sequential minimal optimisation algorithm in the Weka 3-7-11 data mining toolkit [5]. 55 musical features extracted from MIRtoolbox for both the 30/60-second (*N* = 207), and 5-second (*N* = 2990) datasets were used, with the recognition results shown in Table 4.

The RF approach and SVM with linear kernel both performed well, and recognition accuracy using 5-second clips was 1% higher (but not significantly) than for the full excerpts. Although RF using 5-second clips performed best, it still did not reach a satisfactory level. From the confusion matrix, we noticed that classification for the emotion *relaxed* was also collectively confused with *sad*.

4.2.2 Testing

In training, RF gave the best classification accuracy using 5-second clips, and performed time efficiently. Therefore,

[4] https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox/MIRtoolbox1.5Guide

[5] http://www.cs.waikato.ac.nz/ml/weka/

Approaches	Recognition Accuracy	
	30-sec clips	5-sec clips
SVM w/ linear kernel	**39.04%**	40.35%
SVM w/ RBF kernel	28.57%	26.89%
SVM w/ poly kernel	37.62%	29.16%
Random forests	38.57%	**40.75%**

Note. For the training of 5-second clips, the clips from the same song if used in training, were not used for testing. Due to the unbalanced ground truth data for training, the results might be biased.

Table 4. Comparison of classification performance using support vector machines and random forest approaches.

this approach (i.e., RF with 5-second clips) was also applied on the 80 popular musical excerpts. Similar to the data expansion for the training dataset, each audio clip in the testing dataset was also split into 5-second clips (N = 1292). Section 4.3 shows the recognition results in comparison to participants' emotional responses.

4.3 Responses from Participants and the Recognition System

As each excerpt was split into 5-second chunks, each clip was recognised as expressing one emotion. The label with the greatest number of votes from the four emotions was chosen, and the greatest number of votes (consistency) for each excerpt was calculated as well. To compare the responses between outputs from the recognition system and participants for two models of emotion, Pearson's correlation analysis was conducted on the consistency for the 80 musical excerpts.

Table 5 shows that recognition consistency for each excerpt using RF approach is positively correlated with participants' consistency in the categorical ($r(78) = .23$, and $p < .05$) and dimensional models ($r(78) = .36$, and $p < .01$). It tentatively suggests that regardless of the emotion, the consistency of the recognition system is very similar to the consistency of participants' responses.

	Recognition	Categorical
Categorical	.23*	
Dimensional	.36**	.32**

Note. *p <.05, **p <.01.

Table 5. Correlation between the consistency from recognition system using the RF approach and participants' responses.

To explore participants' responses for each emotion, correlation analyses were further conducted on the emotion vote distribution from each excerpt for the recognition system and participants' responses. Table 6 and Table 7 show that no matter which emotion model is used, the emotion vote distribution from the recognition system and participants' responses is highly correlated (i.e., happy, relaxed, and angry). Interestingly, responses for relaxed from the categorical model are also correlated with sad from the recognition system. It suggests that both system and people find it difficult to distinguish between *sadness* and *relaxedness*. The same results could be found in the results for the dimensional model, where significant correlations were shown in the ratings of arousal, whereas only weak correlations were found in the responses of valence (e.g., happy with angry, and relaxed with sad).

		Categorical model			
		Happy	Sad	Relaxed	Angry
Pred.	Happy	.42***	-.33**	-.32**	.13
	Sad	-.16	.18	.33**	-.30**
	Relaxed	-.07	.00	.46***	-.22
	Angry	.07	-.31**	-.39***	.52***

Note. *$p < .05$, **$p < .01$, and ***$p < .001$.

Table 6. Correlation between the responses from recognition system and participants using the categorical model.

		Dimensional model			
		Pos V	Neg V	Pos A	Neg A
Pred.	Pos V	.33**	-.34**	.10	-.12
	Neg V	-.10	-.01	.15	-.16
	Pos A	.22	-.28*	.59***	-.64***
	Neg A	-.02	-.01	-.42***	.44***

Note. *$p < .05$, **$p < .01$, and ***$p < .001$.

Table 7. Correlation between the responses from recognition system and participants using the dimensional model.

Finally, the dominant label(s) from each experiment (recognition system and responses from categorical and dimensional models of emotion) were compared. Considering the same dominant emotion label from both dimensional and categorical model as the ground truth, 32 responses out of 53 (accuracy = 60%) from recognition system were classified correctly. However, if we consider the dominant emotion labels from both categorical and dimensional models, 51 responses out of 80 (accuracy = 64%) are classified correctly.

In spite of the recognition accuracy given by the random forest approach, the incorrect classification results were compared with participants' responses. We found that majority of songs given the incorrect classifications had opposite signs for valence, confusing sad with relaxed and angry with happy. It suggests that compared with arousal, valence is more difficult to recognise. This also agrees with previous studies using regression models [9].

We noticed that if the recognition results were incorrect, it was likely that the emotion of a song itself was ambiguous. Examples for each emotion are provided in Table 8. For example, for the song "Josephine" by *Wu-Tang Clan*, the dominant emotion was chosen as relaxedness in both the categorical and dimensional models, whereas it was recognised as sad by the machine. The distribution, however, shows that 8 clips from the same excerpt were classified as relaxed, and another 10 clips were classified as angry. In addition, participants' responses for both models of emotion were also distributed across four emotions. Similarly, for the song "Blood On The Ground" by *Incubus*, participants all agreed on arousal level, but the responses for valence were ambivalent. Likewise, the recognition result was also influenced by this uncertainty.

Interestingly, we found the song "Anger" by *Skinny Puppy* was recognised as angry by all participants, whereas the recognition system classified it as happy. Possible explanations could be the selection of clips, that some parts are

Title	Recognition					Categorical					Dimensional				
	H	S	R	A	Label	H	S	R	A	Label	PoV	NeV	PoA	NeA	Label
If the Creeks Don't Rise	1	13	9	0	Sad	7	3	5	0	Happy	17	8	16	8	Happy
Loves Requiem	0	2	9	0	Relaxed	0	16	3	0	Sad	1	24	4	23	Sad
Josephine	5	0	8	10	Angry	2	5	6	5	Relaxed	13	10	7	14	Relaxed
Blood On The Ground	3	1	6	1	Relaxed	1	1	3	13	Angry	9	13	24	0	Angry

Note. H - Happy, S - Sad, R - Relaxed and A - Angry.

Table 8. Examples of vote distribution on emotion for the recognition system, categorical and dimensional models.

expressing happiness. It is also reasonable to guess that the emotion perceived is genre-specific (e.g., metal as anger, and pop music as happy) and cultural-dependent. Participants may also be influenced by titles or lyrics.

5. DISCUSSION AND FUTURE STUDIES

In this paper, we presented an empirical study of music and emotion, comparing the results between a music emotion recognition system and participants' responses for two models of emotion. Firstly, we studied the emotional responses for 80 popular musical excerpts in the categorical and dimensional models of emotion. The analysis showed similar responses for both emotion models. A positive correlation between categorical and dimensional models was also found on the rating consistency for each musical excerpt.

A separate 207 musical excerpts were collected from participants for four basic emotion categories (i.e., happy, sad, relaxed, and angry). Our emotion recognition model was trained using support vector machines and random forest classifiers. Two different training datasets were compared, one using the entire 30/60-second audio files and the one using multiple 5-second segments with 2.5-second overlap from the same excerpt. Audio features were extracted using MIRtoolbox. The results showed that the support vector machine with linear kernel and random forest approaches performed best, and the use of 5-second clips increased the classification accuracy by only 1%. In addition, the recognition system did not classify emotions well for the emotions sadness and relaxedness. One of the possible reasons for the low accuracy of music emotion recognition systems could be the user-suggested dataset, which was mixed with both perceived and induced emotion. Another explanation could be the subjective nature of music emotion perception.

Finally, the time-efficient random forest with 5-second clips approach was applied on the 80 musical excerpts for testing. The analysis showed that responses from the recognition system were highly correlated with participants' responses for the categorical and dimensional models. Moreover, the distribution of responses for each emotion was also highly correlated. However, significant correlations between sadness and relaxedness in the categorical model suggest that listeners and emotion recognition systems have difficulty distinguishing valence (positive and negative emotions). Similarly, strong correlations were found for responses of arousal, whereas only weak correlations were shown in responses for valence. The comparison of emotion distribution also indicates that the performance of music emotion recognition systems is similar to participants' emotional responses for the two models of emotion. Additionally, the prediction accuracy is higher for songs where participants agreed more. This suggests that only strongly consistent emotional responses can be predicted by the music emotion recognition systems.

Due to the dynamic nature of music, emotions may vary over time and the emotion recognition accuracy may also be affected by the selection of clips. More importantly, music emotion involves complex interactions between the listener, the music, and the situation. The perception of music is most likely influenced by individual differences such as age, music skills, culture, and music preference [35–38]. The experience of emotions may also vary according to various situational contexts. Therefore, our future work is to incorporate emotion into the design of a subjective music recommendation system, and also to study the influence of situational contexts and individual differences such as culture in the emotion perception of music.

Acknowledgments

We are very grateful to all the participants, and reviewers for their valuable suggestions. We would like to thank the China Scholarship Council (CSC) for financial support and the Centre for Digital Music (C4DM).

6. REFERENCES

[1] P. N. Juslin and P. Laukka, "Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening," *Journal of New Music Research*, vol. 33, no. 3, pp. 217–238, 2004.

[2] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, "Music emotion recognition: A state of the art review," in *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 255–266, Utrecht, Netherlands, 2010.

[3] T. Eerola and J. K. Vuoskoski, "A review of music and emotion studies: Approaches, emotion models, and stimuli," *Music Perception: An Interdisciplinary Journal*, vol. 30, no. 3, pp. 307–340, 2013.

[4] Y. Song, S. Dixon, and M. T. Pearce, "A survey of music recommendation systems and future perspectives," in *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR)*, pp. 395–410, London, UK, 2012.

[5] Y.-H. Yang and H. H. Chen, "Machine recognition of music emotion," *ACM Transactions on Intelligent Systems and Technology*, vol. 3, no. 3, pp. 1–30, 2012.

[6] Y. Song, S. Dixon, and M. T. Pearce, "Evaluation of musical features for emotion classification," in *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 523–528, Porto, Portugal, 2012.

[7] P. Saari, T. Eerola, and O. Lartillot, "Generalizability and simplicity as criteria in feature selection: Application to mood classification in music," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1802–1812, 2011.

[8] T. Eerola, "Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres," *Journal of New Music Research*, vol. 40, no. 4, pp. 349–366, 2011.

[9] Y. Yang, Y. Lin, Y. Su, and H. H. Chen, "A regression approach to music emotion recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 2, pp. 448–457, 2008.

[10] E. M. Schmidt and Y. E. Kim, "Learning emotion-based acoustic features with deep belief networks," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 65–68, New Paltz, New York, USA, 2011.

[11] E. M. Schmidt, J. Scott, and Y. E. Kim, "Feature learning in dynamic environments: Modeling the acoustic structure of musical emotion," in *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 325–330, Porto, Portugal, 2012.

[12] P. N. Juslin and D. Västfjäll, "Emotional responses to music: The need to consider underlying mechanisms," *The Behavioral and Brain Sciences*, vol. 31, no. 5, pp. 559–621, 2008.

[13] T. Eerola and J. K. Vuoskoski, "A comparison of the discrete and dimensional models of emotion in music," *Psychology of Music*, vol. 39, no. 1, pp. 18–49, 2010.

[14] A. Gabrielsson, "Emotion perceived and emotion felt: Same or different?" *Musicae Scientiae*, vol. 5, no. 1, pp. 123–147, 2002.

[15] K. Kosta, Y. Song, G. Fazekas, and M. B. Sandler, "A study of cultural dependence of perceived mood in Greek music," in *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 317–322, Curitiba, Brazil, 2013.

[16] X. Hu and J. H. Lee, "A cross-cultural study of music mood perception between American and Chinese listeners," in *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 535–540, Porto, Portugal, 2012.

[17] E. Bigand, S. Vieillard, F. Madurell, J. Marozeau, and A. Dacquet, "Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts," *Cognition & Emotion*, vol. 19, no. 8, pp. 1113–1139, 2005.

[18] K. Kallinen and N. Ravaja, "Emotion perceived and emotion felt: Same and different," *Musicae Scientiae*, vol. 10, no. 2, pp. 191–213, 2006.

[19] Y. Song, S. Dixon, M. T. Pearce, and A. R. Halpern, "Perceived and induced emotion responses to popular music: Categorical and dimensional models," *to appear in Music Perception*, pp. 1–46, 2015.

[20] S. McAdams, B. W. Vines, S. Vieillard, B. K. Smith, and R. Reynolds, "Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting," *Music Perception: An Interdisciplinary Journal*, vol. 22, no. 2, pp. 297–350, 2004.

[21] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evoked by the sound of music: Characterization, classification, and measurement," *Emotion*, vol. 8, no. 4, pp. 494–521, 2008.

[22] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[23] A. Huq, J. P. Bello, and R. Rowe, "Automated music emotion recognition: A systematic evaluation," *Journal of New Music Research*, vol. 39, no. 3, pp. 227–244, 2010.

[24] P. Ekman, "An argument for basic emotions," *Cognition & Emotion*, vol. 6, no. 3/4, pp. 169–200, 1992.

[25] C. Laurier, J. Grivolla, and P. Herrera, "Multimodal music mood classification using audio and lyrics," in *International Conference on Machine Learning and Applications (ICMPLA)*, pp. 1–6, San Diego, California, USA, 2008.

[26] T. Eerola, O. Lartillot, and P. Toiviainen, "Prediction of multidimensional emotional ratings in music from audio using multivariate regression models," in *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 621–626, Kobe, Japan, 2009.

[27] C. Mak, T. Lee, S. Senapati, Y.-T. Yeung, and W.-K. Lam, "Similarity measures for Chinese pop music based on low-level audio signal attributes," in *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 513–518, Utrecht, Netherlands, 2010.

[28] E. M. Schmidt and Y. E. Kim, "Prediction of time-varying musical mood distributions from audio," in *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 465–470, Utrecht, Netherlands, 2010.

[29] Y.-H. Yang and H. H. Chen, "Prediction of the distribution of perceived music emotions using discrete samples," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2184–2196, 2011.

[30] P. Lamere, "Social tagging and music information retrieval," *Journal of New Music Research*, vol. 37, no. 2, pp. 101–114, 2008.

[31] Y. Song, S. Dixon, M. T. Pearce, and G. Fazekas, "Using tags to select stimuli in the study of music and emotion," in *Proceedings of the 3rd International Conference on Music & Emotion (ICME)*, Jyväskylä, Finland, 2013.

[32] Y. Song, S. Dixon, M. T. Pearce, and A. R. Halpern, "Do online social tags predict perceived or induced emotional responses to music?" in *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 89–94, Curitiba, Brazil, 2013.

[33] I. Peretz, "Listen to the brain: A biological perspective on musical emotions," in *Music and Emotion: Theory and Research*, P. N. Juslin and J. A. Sloboda, Eds. New York, NY, USA: Oxford University Press, pp. 105–134, 2001.

[34] O. Lartillot and P. Toiviainen, "MIR in Matlab (II): A toolbox for musical feature extraction from audio," in *Proceedings of the 8th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 237–244, Vienna, Austria, 2007.

[35] C. Z. Malatesta and M. Kalnok, "Emotional experience in younger and older adults," *Journal of Gerontology*, vol. 39, no. 3, pp. 301–308, 1984.

[36] P. J. Rentfrow and S. D. Gosling, "The Do Re Mi's of everyday life: The structure and personality correlates of music preferences," *Journal of Personality and Social Psychology*, vol. 84, no. 6, pp. 1236–1256, 2003.

[37] M. Shiota, D. Keltner, and O. John, "Positive emotion dispositions differentially associated with Big Five personality and attachment style," *The Journal of Positive Psychology*, vol. 1, no. 2, pp. 61–71, 2006.

[38] D. L. Novak and M. Mather, "Aging and variety seeking," *Psychology and Aging*, vol. 22, no. 4, pp. 728–737, 2007.